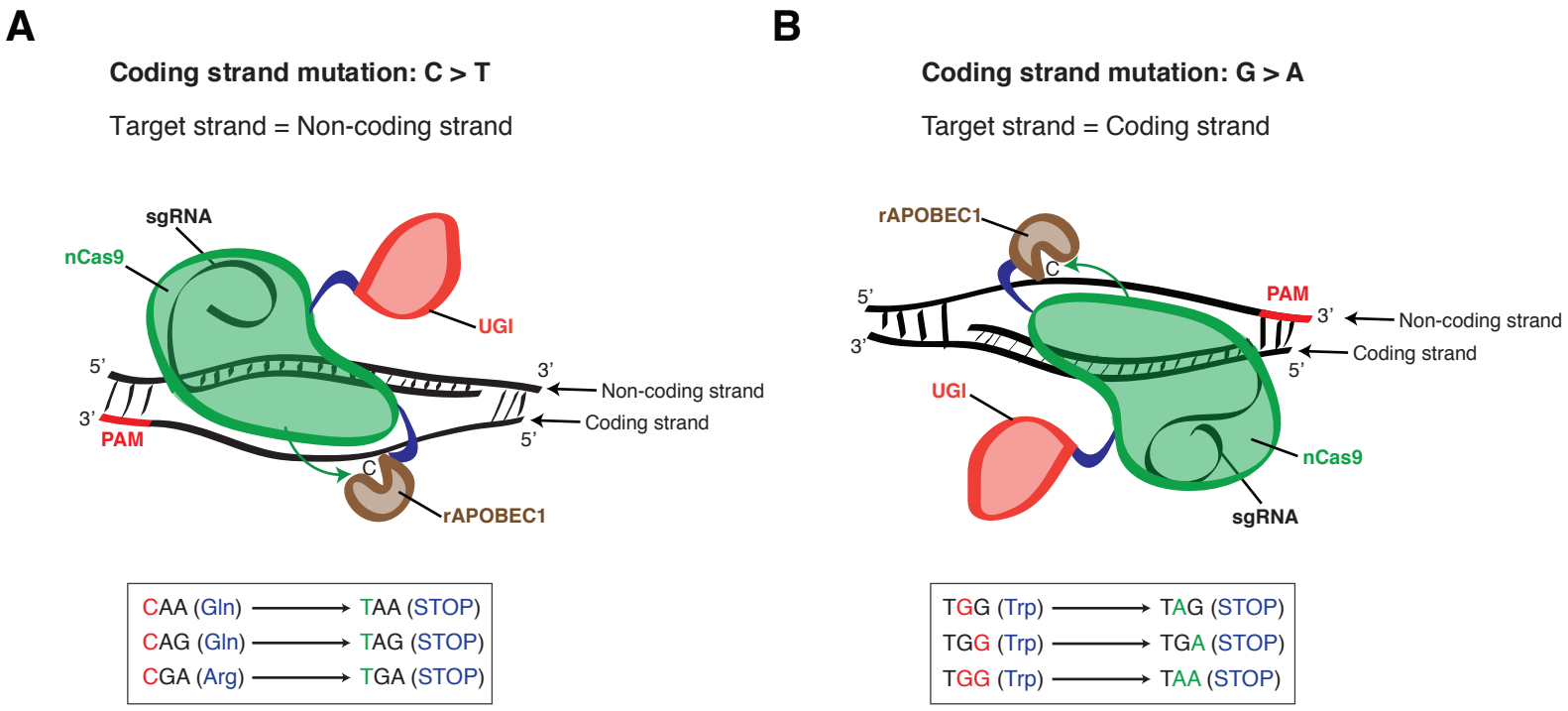
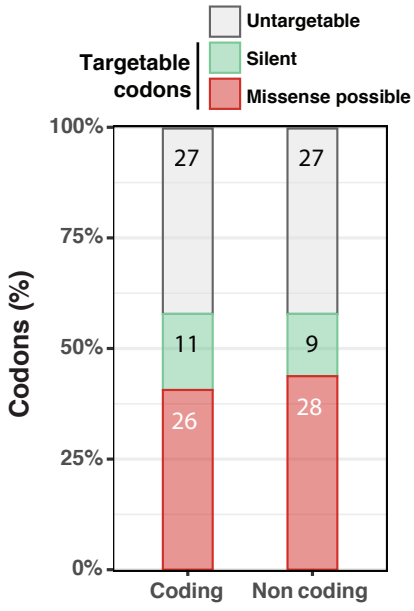
SUPPLEMENTAL INFORMATION

CRISPR-mediated base editing enables efficient disruption of eukaryotic genes through induction of STOP codons (iSTOP)

Pierre Billon, Eric E. Bryant, Sarah A. Joseph, Tarun S. Nambiar, Samuel B. Hayward, Rodney Rothstein and Alberto Ciccia

Supplemental Information includes 5 Supplemental Figures





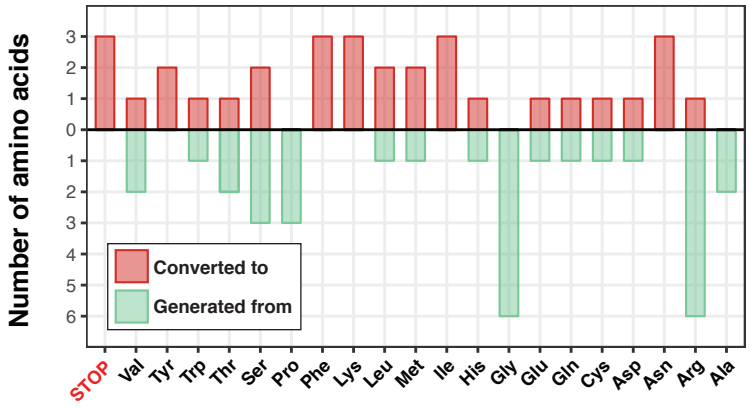


Figure S1 (Related to Figure 1)

Figure S1 (Related to Figure 1). Repertoire of amino acid substitutions generated by CRISPR-mediated base editing on coding or non-coding strands

(A) Schematic representation of the BE3 CRISPR-dependent base editor, consisting of rAPOBEC1 (brown), nCas9 (green) and UGI (red), in the process of targeting a cytosine located on the coding strand. By converting C to T in CAA, CAG and CGA triplets on the coding strand, BE3 generates STOP codons.

(B) Schematic representation of BE3 targeting a cytosine located on the non-coding strand. This process results in G>A transitions on the coding strand. By converting G to A in TGG triplets on the coding strand, BE3 generates STOP codons.

(C) Percentage and number of targetable and untargetable codons and type of mutations generated on the 64 codons by targeting cytidine deaminase-dependent CRISPR base editors on either coding or non-coding strands. Codons untargetable by CRISPR base editors are indicated in white. Codons that when targeted cause silent or possible missense mutations are indicated in green and orange, respectively.

(D) Number of amino acids that can be converted into other amino acids/STOP codons (orange) or generated from different amino acids (green) by cytidine deaminase-dependent CRISPR base editors, as shown in Figure 1B. Ala = Alanine, Arg = Arginine, Asn = Asparagine, Asp = Aspartic acid, Cys = Cysteine, Gln = Glutamine, Glu = Glutamic acid, Gly = Glycine, His = Histidine, Ile = Isoleucine, Met = Methionine, Leu = Leucine, Lys = Lysine, Phe = Phenylalanine, Pro = Proline, Ser = Serine, Thr = Threonine, Trp = Tryptophan, Tyr = Tyrosine, Val = Valine and STOP = STOP codon. See also Table S1.

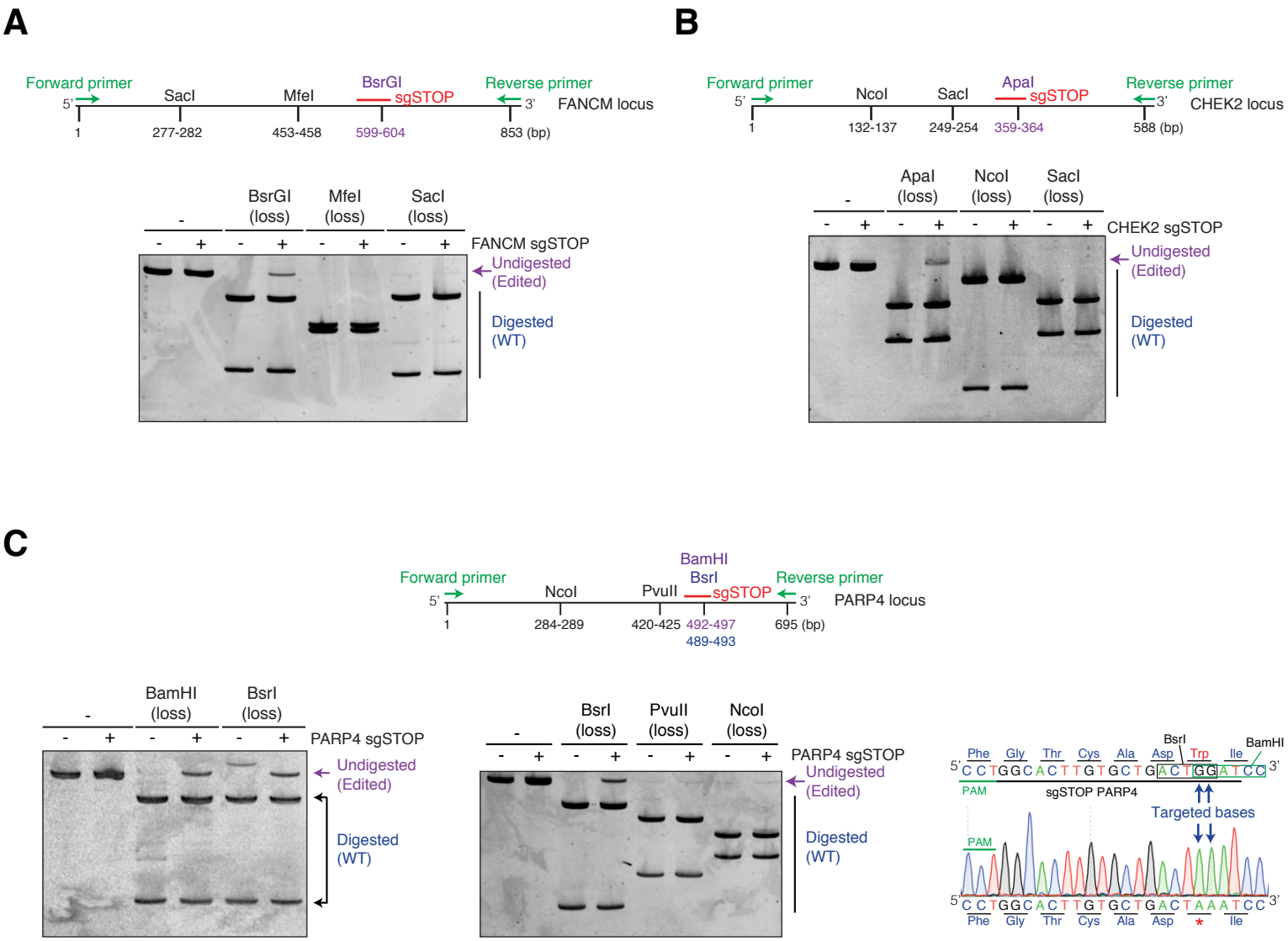


Figure S2 (Related to Figure 2)

Figure S2 (Related to Figure 2). Specificity of the RFLP assay utilized to detect iSTOP-mediated editing

(A) BsrGI-, MfeI- and SacI-mediated digestion of PCR products of the *FANCM* locus targeted with an sgSTOP that edits a BsrGI restriction site. A schematic map of the *FANCM* locus is indicated above.

(B) ApaI-, NcoI- and SacI-mediated digestion of PCR products of the *CHEK2* locus targeted with an sgSTOP that edits an ApaI restriction site. A schematic map of the *CHEK2* locus is indicated above.

(C) BamHI-, BsrI-, PvuII- and NcoI-mediated digestion of PCR products of the *PARP4* locus targeted with an sgSTOP that edits BamHI and BsrI restriction sites. A schematic map of the *PARP4* locus is indicated above. One sequencing profile representative of 4 sequences of *PARP4* amplicons refractory to BsrI digestion is shown on the right inside.

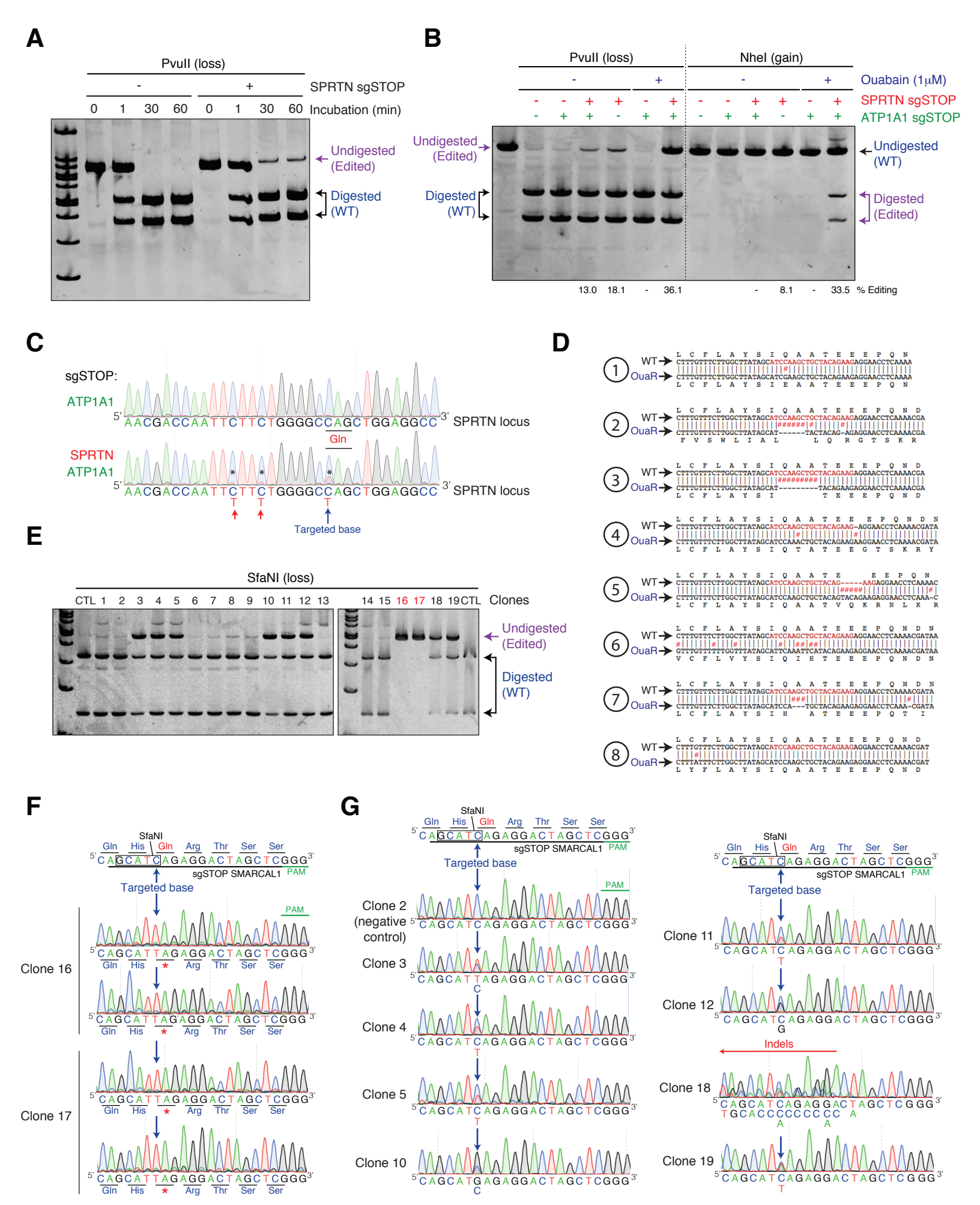


Figure S3 (Related to Figure 3)

Figure S3 (Related to Figure 3). Co-selection strategy to enrich for iSTOP-mediated editing and detection by RFLP assay

(A) PvuII restriction digest of PCR products of the *SPRTN* locus targeted with an sgSTOP that edits a PvuII restriction site. The reaction was terminated after 1, 30 or 60 min and the products of the reaction were run on a polyacrylamide gel.

(B) Digestion of *SPRTN* amplicons from cells transfected with sgSTOPs targeting *SPRTN* and/or *ATP1A1* with or without 1 μM ouabain selection using the restriction enzymes PvuII and NheI. Editing efficiency was monitored by loss of PvuII cutting and gain of NheI cutting, as indicated in Figure 2C.

(C) Sequencing profile of the targeted *SPRTN* locus amplified from cell populations transfected with sgSTOPs targeting *SPRTN* and/or *ATP1A1*. The targeted base that creates a STOP codon (blue arrow) and other targeted bases that generate missense mutations (red arrows) in the *SPRTN* locus are indicated by asterisks (*).

(D) Alignment of sequences of *ATP1A1* alleles from cells targeted with an *ATP1A1* sgSTOP and selected with ouabain. The *ATP1A1* sgSTOP target sequence is represented in red. The symbol "#" indicates a mismatch between the WT genomic sequence (above) and the genomic sequence isolated from ouabain resistant cells (below).

(E) RFLP analysis of the *SMARCAL1* locus from 19 single cell clones derived from HEK-293T transfected with sgSTOPs targeting *SMARCAL1* and/or *ATP1A1* and selected with ouabain. *SMARCAL1* amplicons derived from the above clones were digested with SfaNI, as indicated in Figure 3D.

(F) Sequencing profiles of *SMARCAL1* loci from the SMARCAL1 KO clones #16 and #17 shown in (E). The targeted base is indicated by a blue arrow.

(G) Sequencing profiles of the *SMARCAL1* locus in SMARCAL1 WT and heterozygous mutant clones shown in (E). The targeted base is indicated by a blue arrow and indels are indicated by a red arrow.

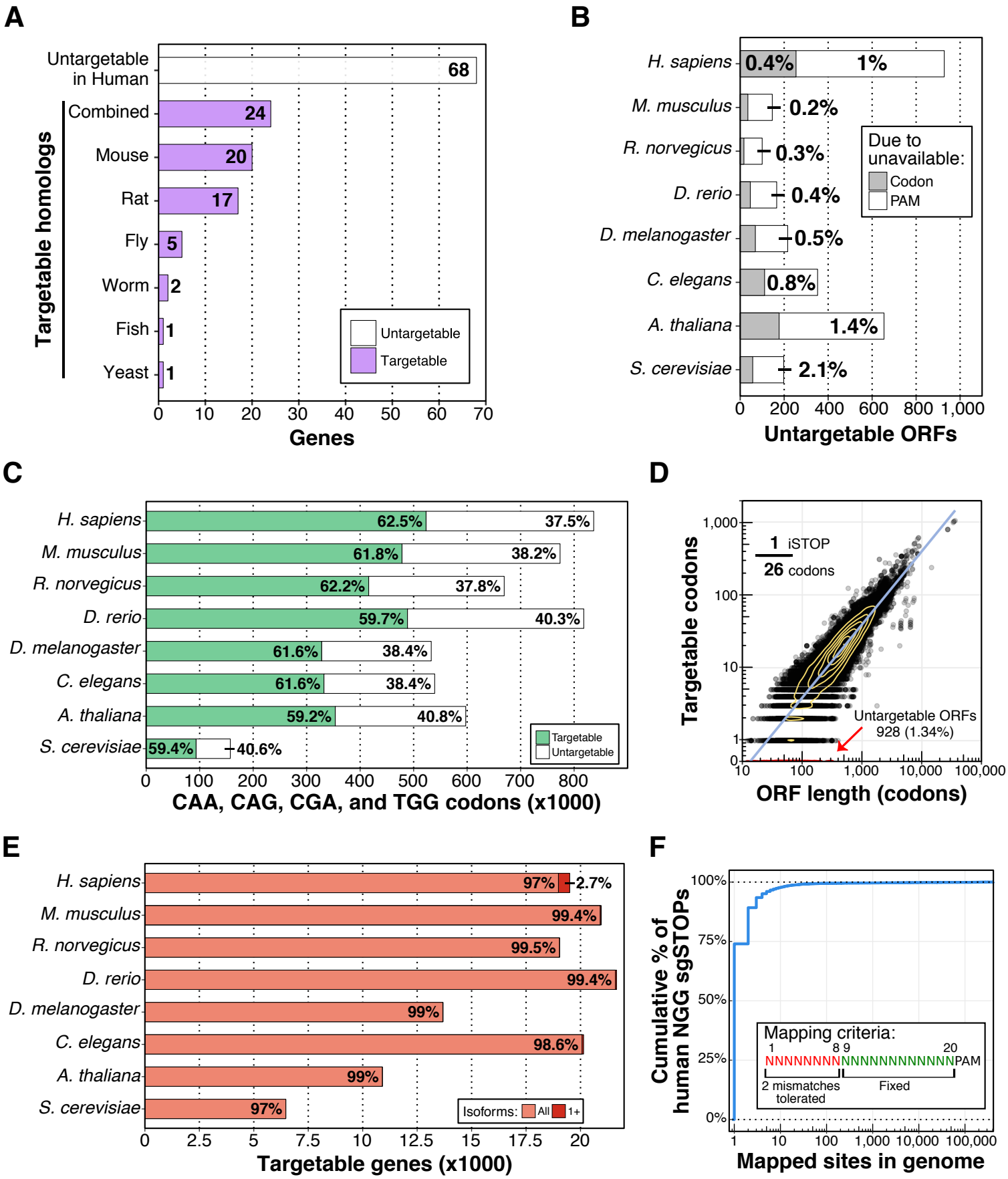


Figure S4 (Related to Figure 4)

Figure S4 (Related to Figure 4). Extended genomic analysis of iSTOP targetable sites in the genomes of 8 eukaryotic species

(A) Human genes untargetable by iSTOP that have targetable orthologs in other eukaryotic species. All homologs reported by Ensembl (www.ensemble.org) were considered. The complete list of untargetable human genes is available in Table S2.

(B) Number of untargetable ORFs in all species considered in this study. Percentage of all ORFs that are untargetable due to an unavailable PAM are indicated in black text.

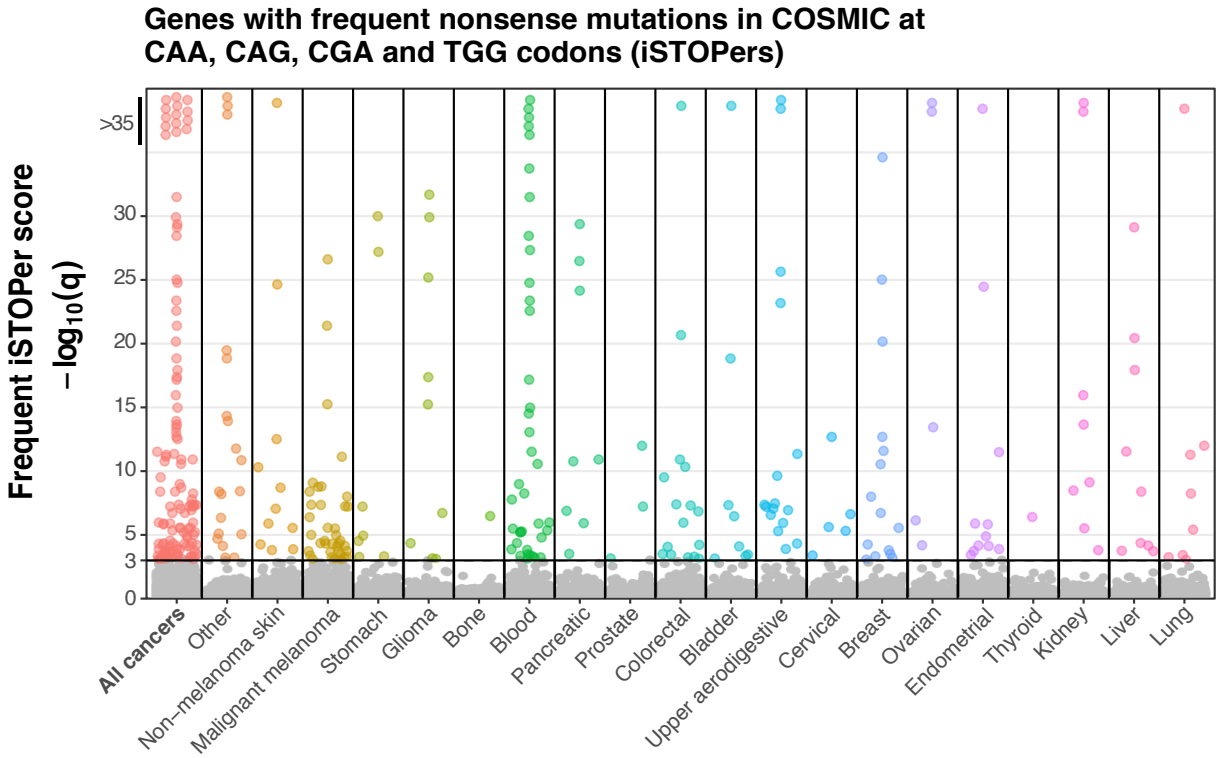
(C) Number of all CAA, CAG, CGA and TGG codons in each species, and whether they are targetable with iSTOP. Percentage of each category is annotated on each bar.

(D) Number of iSTOP targetable codons per ORF length in the human genome.

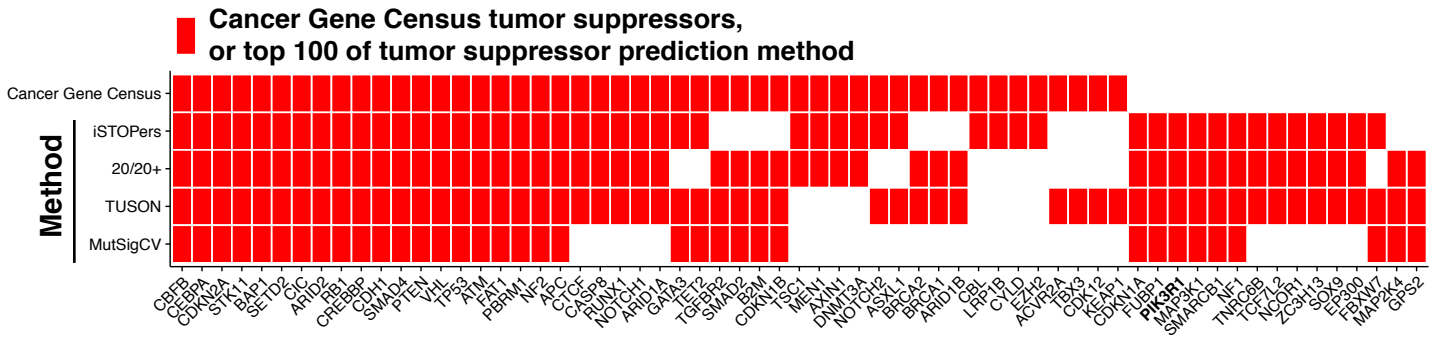
(E) Number of genes that are targetable in all or at least one (1+) isoform. Percentage of total number of genes considered is annotated in text on each bar.

(F) Distribution of the number of mapped sites in the human genome for NGG sgSTOPs. Each NGG sgSTOP was mapped to all matching locations in the genome allowing up to two mismatches outside of the guide's seed sequence (positions 1 through 8). Each guide is expected to map once in the genome. More than 1 mapped site indicates potential for off-target binding.

A



B



C

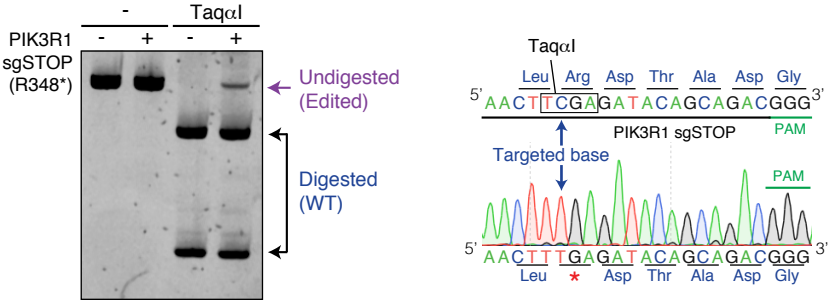


Figure S5 (Related to Figure 5)

Figure S5 (Related to Figure 5). Analysis of genes with frequent nonsense mutations at CAA, CAG, CGA and TGG codons in cancer and modeling of a recurrent cancer-associated nonsense mutation by iSTOP

(A) Genes with frequent nonsense mutations in COSMIC observed at CAA, CAG, CGA and TGG codons (iSTOPers). The frequent iSTOPer $-\log_{10}(q)$ score is derived from an FDR adjusted one-tailed binomial test (STAR Methods, Analysis of frequent iSTOPers). "All cancers" is the highest score observed across all cancer types.

(B) Comparison to tumor suppressor annotation and prediction methods. The top 100 iSTOPers from "All cancers" were compared to Tumor Suppressor Genes (TSGs) annotated by the Cancer Gene Census, the top 100 TSGs from the 20/20+ prediction method, the top 100 TSGs from the TUSON prediction method and the top 100 significantly mutated genes from the MutSigCV method. Genes included in this figure were either annotated as a TSG by the Cancer Gene Census, or were considered a TSG by at least 3 of the 4 TSG prediction methods. The complete list of iSTOPers ($q < 0.001$) is available in Table S4.

(C) TaqαI-mediated digestion of PCR products of the *PIK3R1* locus targeted with an sgSTOP that edits a TaqαI restriction site to generate the cancer associated nonsense mutation R348*. One sequencing profile representative of 4 sequences of *PIK3R1* amplicons refractory to TaqαI digestion is shown on the right inside.